DOCUMENT RESUME

ED 191 445 IR 008 677

AUTHOR Anderson-Selander, Sandy E.
TITLE A Syntax for the Machine Readable Description of Numeric Data.
PUB DATE 31 Dec 79
NOTE 34p.: Paper presented at the National Online Information Meeting (New York, NY, March 25-27, 1980). Not available in paper copy due to poor print quality of the original.

EDRS PRICE MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS *Data Bases; *Data Processing; Information Retrieval; Information Storage; *Methods; *Online Systems; *Syntax
IDENTIFIERS *Numeric Data

ABSTRACT

A syntax methodology provides a descriptive mechanism to enable machine readable storage and retrieval of numeric data contained in bibliographic documents. The body of this paper treats the nature of "the syntax," "the data element," and the language by which this syntax is expressed. A discussion of the methodology illustrates a possible data browsing technique in an heuristic online environment. Appendices provide (1) examples of the original numeric data and the transcriptions of those data into the syntax notation; (2) representation in the relational calculus of the technical formulation for the essence of the syntax; and (3) 13 rules of order that must be followed when attempting to transcribe the numeric data into the syntax. (Author/RAA)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED191445

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

BEST AVAILABLE COPY

A

SYNTAX FOR THE MACHINE READABLE DESCRIPTION

OF

NUMERIC DATA

BEST AVAILABLE COPY

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Sandy E.
Anderson-Selander
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Sandy E. Anderson

31 December 1979

I200877

429

TABLE OF CONTENTS

		Page
I.	INTRODUCTION	1
II.	BACKGROUND	2
III.	A SYNTAX FOR THE DESCRIPTION OF NUMERIC DATA	4
	A. Building the Syntax	4
	B. The Data Element	8
	C. A Language for Syntax Expression	9
IV.	APPLICATIONS FOR BROWSING	11
V.	References	14
	APPENDICES	15
	Appendix 1 - Numeric Data Descriptions	16
	Appendix 2 - Syntax and General Rules for Logical Data Definitions	24
	Appendix 3 - Logical Data Declarations	27

I. INTRODUCTION

This paper describes a methodology by which numeric data, contained within bibliographic documents, may be precisely described in order to enable machine-readible storage and retrieval. This paper represents work which was funded by the Department of Energy-Office of Technical Information, through the Oak Ridge National Laboratory-Data Management and Analysis Group, during the period June 1978-1979.

The body of this paper is partitioned to treat the nature of "the syntax," "the data element," and the language by which this syntax is expressed, as three separate discussions (Subsections A, B, C within Section III respectively). Section IV discusses how this methodology might be applied to a heuristic, on-line environment to facilitate "data base browsing." The appendices (1) provide examples of the original numeric data and the transcription of that data into the syntax notation, (2) present in the relational calculus the technical formulation for the essence of the syntax, and (3) enumerate thirteen rules of order which must be followed when attempting to transcribe the numeric data into the syntax.

II. BACKGROUND

The DOE-Office of Technical Information (DOE-TIC) is charged with, among other things, providing both Congress and the American public with energy related information within the public domain. Although there already existed a comprehensive bibliographic resource, the RECON data base, for the DOE and its contractors' use, attention was now turned toward developing a mechanism by which numeric data, contained within the bibliographic document itself, could be described, stored and retrieved. It was felt that the ability to "browse" through a numeric "abstract" of sorts would cut down the amount of time it would take a researcher to find desired numeric information. While the RECON data base served as an adequate bibliographic resource, there remained to be developed a syntax by which indexers/ abstractors could precisely and consistently record the existence of numeric information-as well.

In June of 1978 the DOE-OTI authorized funding for researching a methodology for the precise description of numeric data. The two phased objective was to first define a syntax for numeric data description; and subsequently to test it against some documents containing numeric data. By the end of 1978, working in conjunction with Arthur D. Little, Inc., Cambridge, Massachusetts, a methodology was developed by which one could describe the essence of numeric information contained within a document. The author's work for this paper was conducted while she was employed with Evaluation Research Corporation, Vienna, Virginia. This paper describes the structure of the syntax; how each piece is defined and used in conjunction with other syntax elements to effect a clear and consistent description of numeric data. The sample numeric data found in Appendix 1 may be presumed to have been taken from Chapter 2 of the "Regional Transportation Energy Conservation Data Book (R.T.E.C.D.B.)," published by the Data Management and Analysis

Group of the Oak Ridge National Laboratory, Oak Ridge, Tennessee, unless otherwise noted.

III. A SYNTAX FOR THE DESCRIPTION OF NUMERIC DATA

A. Building the Syntax

This section describes the formulation of a syntax for numeric data description and defines a "language" into which the syntax may be applied to facilitate the retrieval of numeric data.

In order to clearly understand the basis of the syntax proposed herein, it may be helpful to briefly touch upon the underlying basis of relational data bases, as this syntax has been designed to be used within a relational data base environment. Locked into the "meaningfulness" of a statistic, within a table, graph or chart, is a unique relationship between the intersecting column and row headings. This unique intersection for each value results in the value taking on a distinct "identity," each facet of which we refer to as the "inherent attribute," or "inherent dimension" of the numeric value. When these attributes are linked together and associated with a distinct numeric value, the value gains a "meaningfulness" which helps to keep it distinct from all other values in the collection of values. Thus, the numeric value "697,652,831" takes on an identity when it is associated with the inherent dimensions, "Gasoline Consumption," "Virginia," and "1975." This value may now be differentiated from the value "523,008,937" which may be associated with the inherent dimensions "Coal Production," "Domestic-U.S.," "1978," "Short Tons".

When building a syntax for the description of numeric data, care must be exercised to insure against misinterpretation or misstatement of the data. The following example illustrates the pitfalls of not accommodating the full contextual properties of numeric data.

Consider the example "fuel consumption by state by year." The particular values of "1000 for Pennsylvania in 1977" may very well

be an adequate definition for its original purpose, but many aspects of its meaning remain unspecified. Each attribute of this data element may be further specified by refining the definitions of each component (domains) and their specific roles in the data definition. For example:

Domains:	Roles in the relation:
States,	Place where the fuel was consumed,
Years,	Year when fuel was consumed,
Quantities,	Amount of fuel consumed in thousands of gallons.

However, even the explanation fuel consumption did not specify gasoline (or diesel fuel or heating fuel) in automobiles (or trucks or houses) in fiscal years (or calendar years). Although it may be impossible to account for all the attributes associated with the contextual meaning of a numeric value, this syntax allows for those, most significant attributes to be described.

Having agreed thus far in this discussion that: 1) numeric values gain their identities from association with "inherent attributes," and 2) the contextual meaningfulness of these inherent attributes must be defined precisely in order to insure against misstatement or misinterpretation of the values, attention may now be directed to the creation of the syntax itself.

Of fundamental importance to the creation of the syntax is the notion of a data value, for it is upon this component that all other building blocks are developed.* A "data value" may be defined as a symbol which identifies a unit of descriptive information having a

*the terms data value, N-tuple, domain, and dimension are acknowledged to have been developed by E.F. Codd. For a detailed discussion of their development the reader may refer to the references listed in Section V.

unique meaning. Massachusetts, Virginia, New York, etc. are all examples of data values. Correspondingly, an "N-tuple" is understood to represent a collection of N data values, one from each of N not necessarily disjoint domains. The collection >1000, Massachusetts, 1978< is a three-tuple of values from domains "quantity," "state" and "year." "Domain," precisely defined may be considered a named class of data values which identify related objects. A domain may be specified by giving its:

- name
- description of legal values and their meaning
- ordering (meaning of ≤)
- authority for allowing items in the domain.

For example, "States of the U.S." is a domain, one of whose values is "Massachusetts."

A group of related domains, each having a single measurable objective, but differing in units of measure, precision or representation, is labeled a "dimension." Thus, the related domains year, month, and day all refer to the dimension "time." Correspondingly, the term "metric" is reserved for a domain (or dimension) whose values are chosen from a continuous (possibly finite) range and "distance" would be one such metric domain. "Nominal" domains contrastingly represent domains (or dimensions) whose values are chosen from a discrete (possibly infinite) set of (usually literal) values, and "state" would be one such nominal dimension.

Next we come to the notion of a "relation." A relation is a set of N-tuples defined over the same collection of domains. A relation may be specified intentionally by designating: (1) the domains involved, and (2) the rule for determining which tuples defined over those domains are in the relation. Alternately, a

relation may be specified extensionally by listing all its tuples. Or a relation may be specified functionally, as an interpretation of an expression in the relational calculus (see Appendix 3).

In this report a relation is usually specified by giving any one of the following:

- its name,
- the domains and their roles,
- a narrative description of the meaning of a tuple and the roles of its domain values,
- the condition for including tuples in the relation usually as,
  a. an instance of a relationship, or
  b. a functional dependency.

Example: A relation QSY is defined over domains quantity, state, and year having tuples such as:
<1000, Massachusetts, 1978>
<4000, New York, 1978>
<4150, New York, 1978>.

When domains are combined to form a relation, it usually results that one domain may become dependent upon the conditions expressed by the others. In this case, a functional dependency results. A domain "A" of relation R is said to be functionally dependent upon domains B, C, (etc.)...of R, if two tuples having identical values for domains B, C, (etc)...would consistently have the same value in domain A.

For example, in the QSY relation above, the "quantity" domain is defined to be functionally dependent upon the "state" and "year" domains.

Whenever meaningfulness is meant to be derived from a string of

values forming an N-tuple, a relationship is said to exist. A relationship constitutes a rule for the assignment of meaning to an N-tuple. When a relationship is composed of more than one domain, the meaning of each domain is its role in the relationship. To illustrate these concepts, let us consider the relationship "fuel consumed by automobiles in a state during a calendar year," covering three domains "quantity," "state," and "year." The role of quantity in this relationship is "how much;" the role of state is "where;" and the role of year is "when."

B. The Data Element

Having examined each piece of our syntax individually, they may now be used to build an understanding of the notion of a data element. A "data element" is a named unit of descriptive information having unique values from particular domains and a unique contextual meaning. There are two basic concepts here:

- a unit of descriptive information whose value is chosen from a particular domain
- a unique contextual meaning associated with that value.

These concepts, when transformed into the syntax of data element description, yield the following composition for every data element.

descriptive unit	ELEMENT: name-of-domain
	OF: role-in-relationship
	IN: measurement-in-relationship

followed by one or more identifying clauses:

identifying clauses	BY: domain-range
	FOR: domain EQ: value

and further context-clarification clauses:

contextual clauses	WHERE: existence-condition
	SOURCE: authority

Thus, our earlier example of data element "fuel consumption by state by year" yields the descriptive unit "fuel consumption;" the

identifying clauses "by state" and "by year," and the contextual clause "where consumption is understood to represent gasoline consumption in automobiles only for calendar year 1979."

A more general way of looking at this concept of the data element is to imagine the data element to be a point in an N-dimensional space with one dimension as the locus for the data element value and additional dimensions expressed by the domains for the identifying clauses. As many of these additional dimensions as are necessary to uniquely specify the meaning of the data must be included in the definition. Non-quantitative context is supplied by the WHERE clauses (similar to footnotes to the abstract of a book).

C. A Language for Syntax Expression

This section defines a language by which the syntax may be expressed. These specifications define only the general format for the data description statements, as they are entered into a data base. A more complete specification would also include syntax rules and general rules: A syntax rule supplements the general formats in defining what may be written in a schema. It explains or restricts the usage of elements in the formation of clauses, phrases, and entries. A general rule defines or clarifies the meaning or function of a clause or of its constituent elements. It may also define the relationship between the functions of one clause and other clauses.

The notation used to describe the rules of the syntax is as follows:

- All underlined upper case words are required when the format is used.
- Upper case words which are not underlined are optional words and need not be used.
- Lower case words are generic terms which must be replaced by appropriate names or values.

- The meaning of enclosing a portion of a general format in special symbols as follows is:

```
[a]
[b]   at least no occurence of this component
[c]                    and
      at most one occurrence of this component of the format

{a}
{b}   at least one occurrence of this component
{c}                    and
      at most one occurrence of this component of the format

‖a‖
‖b‖   a least one occurrence of this component
‖c‖                    and
      at most one occurrence of all from within the range
      of possibilities within this component of the format
```

- An ellipsis (that is, '...') indicates repetition is allowed. The portion of the format which may be repeated is determined by the '[' or '{' which logically matches the ']' or '}' to the immediate left of the '...' A comma or 'AND' and one or more spaces must appear between repetitions of the portion of a format delimited by ' ' and ' '.

There are two main sections of a data base description: logical data definitions, and physical data definitions.

- Logical Data Definitions include declarations of the domains of the data base, the hierarchy of dimensions, and the links between data elements of the data base, as well as declarations of the data elements of the data base.
- Physical Data Definitions include declarations of the fields of the data base and access paths to data records of those files, as well as declarations of the fields and records of the data base.

For a detailed presentation of both the logical and physical data declarations transcribed into the language of the syntax, the user is referred to Appendix 3.

IV. APPLICATIONS FOR BROWSING

In Section III, a technical discussion of the fundamental concepts of the proposed syntax for the description of numeric data was presented. In Section IV, the potential for applying this syntax and the impact it might have in terms of user-system interactions are discussed below.

When browsing in a data base, users sometimes come upon additionally relevant numeric data which may only be peripherally related to their primary interests. In this case, the user should be able to scan this numeric data to decide if it is suited to his needs or if it requires some modification (e.g. of the values) to become more useful. The syntax for numeric data description proposed herein, provides a means by which numeric data elements may be specified and retrieved in fulfillment of such a need. This proposed syntax would enable a computer to identify all related forms of the subject area queried. The dependent domain for each data group may be displayed in conjunction with all the independent domains with which it is known to be associated within the data base. For example, consider a query "coal production." In this instance a user would want to have displayed all data group headings in which "coal production" acting as a dependent domain, would be stored internally in conjunction with all various combinations or other independent domains. The total number of potentially relevant, retrievable relations would vary depending upon how the originators of the numeric data organized the data; and, e.g., in the case of tables, labeled the data's column and row headings.

Admittedly, it would be impossible for a data base administrator to control for inconsistency in the assignment of numeric data terminology* Therefore, care must be taken to refrain from becoming

*The assignment of terminology is presumed to occur by the authors or statistical compilers at the time the document is created.

misled by the data group headings. Once the system has responded with the number and name of all "coal production" relations, the user may then begin to select the relation(s) which best meet the need. Homing in on any specific table, bargraph, or histogram would result from the user choosing one relation of "coal production" and asking to see displayed either (1) the bibliographic record first, and then the numeric records representing each significant "numeric document," or (2) going directly to a display of each numeric record. In a numeric data base, option 2 would be the system default. Thus, a typical scenario might be as follows:

SYSTEM: Welcome to the fossil fuels data base. Please enter the desired data element.

USER: COAL PRODUCTION

SYSTEM: The term coal production is represented in the fossil fuels data base as follows:

		CITS
(1)	COAL PRODUCTION BY: STATE BY YEAR EQ 1965-1975	43
(2)	COAL PRODUCTION BY: U.S. DOMESTIC FOR: YR EQ: 1977	6
(3)	COAL PRODUCTION BY: U.S. REGION FOR: YR EQ: 1970	14
(4)	COAL PRODUCTION BY: WORLD WIDE FOR: YEARS, PROJECTED EQ: 1980-2000	5

SYSTEM: Please enter the desired data element:

USER: 2

SYSTEM: Coal Production BY: U.S. Domestic FOR: YR EQ: 1977 Yields 6 items, do you wish numeric record display only?

USER: Y

SYSTEM: Do you wish output to line printer?

Do you wish CRT display?

Do you wish batch output?

Please specify desired option.

USER: Display

SYSTEM: The following numeric records were found: (6 Numeric Descriptions Follow)

Through the scrolling process, a user could page through the numeric descriptions (representations) of the major numeric data within e.g., a technical report, and in that way make a more educated judgment about the suitability of the material. While this methodology brings us just short of reporting the actual numeric value, it brings us one step closer to the day when machines will be able to tell us exactly "how much" or "how many" or "where" and "when." It is conceivable that this methodology, applied within the "relational data-base" environment espoused by Codd in his writings (cited below) would offer great promise for the future development of numeric data bases.

V. REFERENCES

Codd, E.F. "A Relational Model For Large Shared Data Banks," Communications of the ACM, Volume 13 #6, June '70 pgs. 377-381.

Codd, E.F. "A Data Base Sublanguage Founded on the Relational Calculus," ACM-SIG FIDET, worksnop on Data Description, Access and Control, 1971 pgs. 35-68.

Codd, E.F. "Further Normalization of the Data Base Relational Model," in Data Base Systems, Courant Computer Science Symposia, Volume 6, Prentice Hall, Englewood Cliffs, New Jersey, May 1971.

Codd, E.F. "Relational Completeness of Data Base SubLanguages," in, Data Base Systems, Courant Computer Science Symposia, Volume 6, Prentice Hall, Englewood Cliffs, New Jersey, May 1971.

APPENDICES

APPENDIX 1

NUMERIC DATA DESCRIPTIONS

AVIATION FUEL SALES

	Aviation gas			Jet fuel			Aviation energy use, 1975	
	1974	1975	Percent change	1974	1975	Percent change	Btu × 10^{12}	Btu × 10^{6} per capita
Region VI								
Arkansas	7,002	7,647	9.21	10,811	12,834	18.71	2.68	1.27
Louisiana	9,379	9,521	1.51	208,505	145,515	-30.21	20.83	5.49
New Mexico	5,177	5,243	1.27	36,678	36,623	-0.15	5.59	4.89
Oklahoma	7,386	7,897	6.92	62,114	62,637	0.84	9.43	3.48
Texas	32,122	33,521	4.35	667,548	718,504	7.63	10.20	8.27
Subtotal	61,066	63,829	4.52	985,656	976,113	-0.96	130.70	6.35
Region VII								
Iowa	6,756	6,935	2.65	21,322	24,279	13.87	4.14	1.44
Kansas	9,799	10,457	6.72	23,444	21,244	-9.38	4.16	1.84
Missouri	11,257	9,369	-16.77	263,567	283,180	7.44	39.39	8.27
Nebraska	5,071	5,424	6.96	30,394	27,614	-9.15	3.40	2.85
Subtotal	32,883	32,185	-2.12	338,727	356,317	5.19	51.09	4.55
Region VIII								
Colorado	8,038	7,875	-2.03	250,077	239,639	-4.17	33.33	13.12
Montana	3,881	3,930	1.26	20,962	22,326	6.51	3.50	4.69
North Dakota	2,580	2,912	12.87	10,078	10,201	1.22	1.74	2.73
South Dakota	2,612	2,693	3.10	7,071	7,150	1.12	1.29	1.91
Utah	3,629	3,777	4.08	40,344	42,498	5.34	6.21	5.16
Wyoming	2,605	2,712	4.11	4,481	3,062	-31.67	0.75	1.99
Subtotal	23,345	23,899	2.37	333,013	324,876	-2.4	46.82	7.57
Region IX								
Arizona	10,213	10,174	-0.38	124,539	126,021	1.19	18.27	8.26
California	92,289	70,515	-23.59	1,640,869	1,723,628	5.04	241.40	11.39
Hawaii	2,769	3,245	17.19	445,248	441,278	-0.89	59.97	69.10
Nevada	4,637	4,378	-5.59	131,953	143,599	8.83	19.93	33.78
Subtotal	109,908	88,312	-19.64	2,342,609	2,434,526	3.92	339.60	13.66
Region X								
Alaska	14,379	14,447	0.47	218,307	215,598	-1.24	30.90	87.78
Idaho	4,359	4,194	-3.79	10,446	10,441	-0.05	1.93	2.37
Oregon	6,560	6,514	-0.70	73,207	67,608	-7.65	9.93	4.34
Washington	9,050	9,547	5.49	340,447	401,002	17.79	55.32	15.60
Subtotal	34,348	34,702	1.03	642,407	694,649	8.13	98.08	14.01
U.S. total	436,472	427,978	-6.22	9,895,796	10,055,222	1.61	1410.52	6.62

[2]National Airport is in Virginia.

Source: Unpublished data furnished by the Federal Aviation Administration.

SAMPLE 1. DATA GROUP - Avaition Fuel Sales

ELEMENT	Sales	OF	Aviation Gas	IN	thousands of gals
ELEMENT	Sales	OF	Jet Fuel	IN	thousands of gals
ELEMENT	Use	OF	Aviation Energy	IN	trillions of BTUs
BY	Region	OF	Sale		
BY	State	OF	Sale		
FOR	Year	EQ	1974-75		
WHERE	Washington, D.C. National Airport is included in Virginia				
SOURCE	R.T.E.C.D,B.		pp.		160-61

Summary of State and Regional Air Activity Measures

	State[a]			Regional[b]		
	High	Low	Mean	High	Low	Mean
Overall activity measures						
Total flight operations, 10^3	10,754 (CA)	168 (WY)	1168	12,712 (IX)	2592 (VIII)	5886
% of U.S. itinerant operations	15.6 (CA)	0.3 (WY)	1.98	19.2 (IX)	4.04 (VIII)	10.0
Itinerant operations/capita	1.6 (AK)	0.08 (MD)	0.18	0.33 (X)	0.10 (II)	.18
General aviation						
Active aircraft, 10^3	21.3 (CA)	0.3 (RI)	3.3	30.9 (V)	6.1 (I)	16.8
Aircraft per 1000 population	11.9 (AK)	0.3 (NY)	0.7	2.0 (X)	0.4 (II)	0.7
Air carrier activity						
Aircraft departures, 10^3	356 (CA)	0 (DE)	88	847 (V)	200 (I)	453
Departures/capita, 10^{-3}	331.6 (AK)	0 (DE)	21.2	41.7 (VIII)	15.7 (II)	21.2
Passenger enplanements, 10^6	19.5 (CA)	0 (DE)	3.8	36.7 (IV)	7.1 (I)	19.5
Enplanements/capita	8.3 (HI)	0 (DE)	0.91	1.5 (VIII)	0.6 (I)	0.91
% Locally generated travel	94 (RI)	26 (CA)	68	84 (II)	52 (VIII)	68
Forecasts						
General aviation growth rate in total operations between 1976 and 1988[c]	82 (MA)	45 (ME)	64	71 (V)	59 (I)	64
Total aircraft operation growth rate between 1975 and 1988[d]	44 (ND)	44 (ME)	69	83 (X)	53 (II)	69

SAMPLE 2

DATA GROUP: Summary of State and Regional Air Activity Measures

ELEMENT	Operations	OF	total	IN	thousands
ELEMENT	Operations	OF	itinerants	IN	per cent
ELEMENT	Operations	OF	itinerants	IN	per capita
ELEMENT	Aircraft	OF	total	IN	thousands
ELEMENT	Aircraft	OF	total	IN	per 1000 pop
ELEMENT	Departures	OF	aircraft	IN	thousands
ELEMENT	Departures	OF	aircraft	IN	per capita(10^3)
ELEMENT	Enplanements	OF	passengers	IN	millions
ELEMENT	Enplanements	OF	passengers	IN	per capita
ELEMENT	Travel	OF	locally generated	IN	per cent
ELEMENT	Operations	OF	general aviation growth	IN	ratio
ELEMENT	Operations	OF	general aviation growth	IN	ratio
BY	State	OF	high activity		
BY	State	OF	low activity		
BY	total state	OF	mean activity		
BY	Region	OF	high activity		
BY	Region	OF	low activity		
BY	total region	OF	mean activity		
FOR	year	OF	Operations	EQ	1975
WHERE	the growth ratios for general aviation is between the years 1976 and 1988 and for total aircraft between the years 1975 and 1988, not including Washington, D.C.				

CERTIFICATED AIR CARRIER ACTIVITY

	Activity by type			Percent of total U.S. air carrier		
	Aircraft departures	Passenger enplanements	Cargo-tons enplaned	Aircraft departures	Passenger enplanements	Cargo-tons enplaned
Region VI						
Arkansas	29,624	622,987	4,674	0.65	0.32	0.13
Louisiana	74,683	2,906,756	26,144	1.65	1.49	0.72
New Mexico	31,447	903,511	6,118	0.69	0.46	0.17
Oklahoma	49,484	1,512,121	19,228	1.09	0.78	0.53
Texas	279,428	13,182,957	187,923	6.17	6.78	5.21
Subtotal	464,666	19,128,332	244,117	10.27	9.83	6.77
Region VII						
Iowa	45,300	973,381	9,560	1.00	0.50	0.26
Kansas	21,937	818,358	5,538	0.48	0.27	0.15
Missouri	155,048	5,003,279	81,311	3.45	5.09	2.25
Nebraska	41,426	994,485	8,619	0.92	0.51	0.24
Subtotal	263,711	8,490,203	105,028	5.85	4.36	2.91
Region VIII						
Colorado	125,506	6,148,436	74,400	2.77	3.17	2.06
Montana	33,528	656,424	3,587	0.74	0.34	0.10
North Dakota	18,541	381,698	2,353	0.41	0.20	0.07
South Dakota	29,851	458,335	3,002	0.66	0.24	0.08
Utah	32,180	1,384,427	15,576	0.71	0.71	0.43
Wyoming	18,463	237,073	1,229	0.41	0.12	0.03
Subtotal	258,069	9,276,393	100,147	5.70	4.77	2.78
Region IX						
Arizona	64,905	2,715,721	21,354	1.43	1.40	0.59
California	356,654	19,453,389	625,101	7.88	10.00	17.33
Hawaii	92,907	7,214,156	127,689	2.05	3.71	3.54
Nevada	56,571	3,447,138	6,186	1.25	1.77	0.17
Subtotal	571,037	32,840,404	780,330	12.62	16.88	21.63
Region X						
Alaska	116,729	1,667,571	90,975	2.58	0.86	2.52
Idaho	19,516	537,065	2,438	0.43	0.28	0.07
Oregon	49,184	1,827,296	25,416	1.09	0.94	0.70
Washington	70,283	3,658,430	150,526	1.55	1.81	4.17
Subtotal	255,712	7,690,362	269,355	5.65	3.95	7.47
Total U.S.	4,525,031	194,538,351	3,607,860	100.0	100.0	100.0

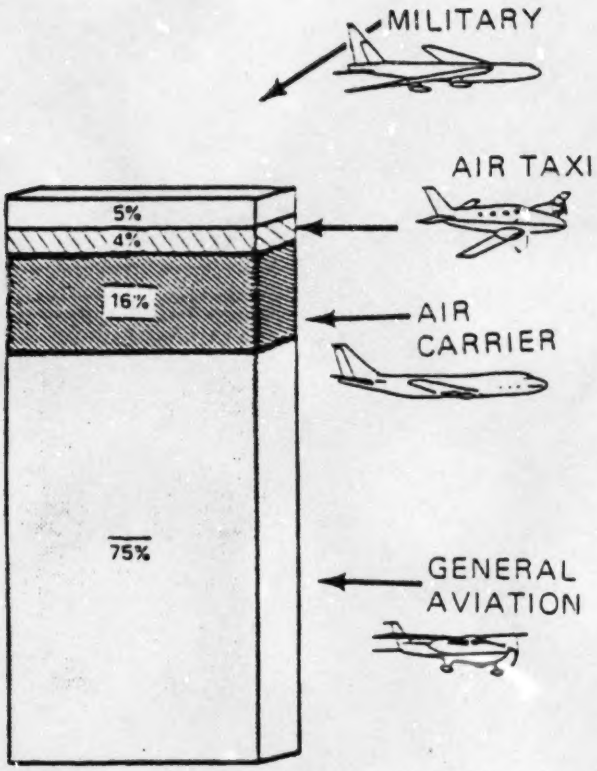
[a]See page 169 for a summary of the table content.

Source: Civil Aeronautics Board and U.S. Department of Transportation, Federal Aviation Administration, Airport Activity Statistics of the Certificated Route Air Carriers, 12 Months Ended December 31, 1975, Washington, D.C., December 1975, Table 2.

SAMPLE 3

DATA GROUP: Certificated Air Carrier Activity

ELEMENT	Departures	OF	Aircraft	IN	count
ELEMENT	Departures	OF	passengers	IN	percent of U.S. total
ELEMENT	Enplanements	OF	passengers	IN	count
ELEMENT	Enplanements	OF	cargo-tons	IN	percent of U.S. total
ELEMENT	Enplanements	OF	cargo-tons	IN	count
ELEMENT	Enplanements	OF	cargo-tons	IN	percent of U.S. total
BY	Region	OF	U.S.		
BY	State	OF	U.S.		
FOR	Year	OF	Activity	EQ	1975
SOURCE	CAB and DOT/FAA Airport Activity Statistics of the Certificated Route Air Carriers, 12 Months Ended December 31, 1975, Washington, D.C., December 1975, Table 2.				



Percent Breakdown of Aircraft Operations
at the U.S. Aggregate Level.

Source: U.S. Department of Transportation, Federal Aviation Administration, Aviation Forecasts FY 1977–1988, Washington, D.C., December 1976, p. 29.

SAMPLE 4

DATA GROUP: Percent Breakdown of Aircraft Operations

ELEMENT	Operations	OF	aircraft	IN	percent
FOR	Aircraft type		EQ		military, air taxi, air carrier, general aviation
FOR	Fiscal year		EQ		1977-1978

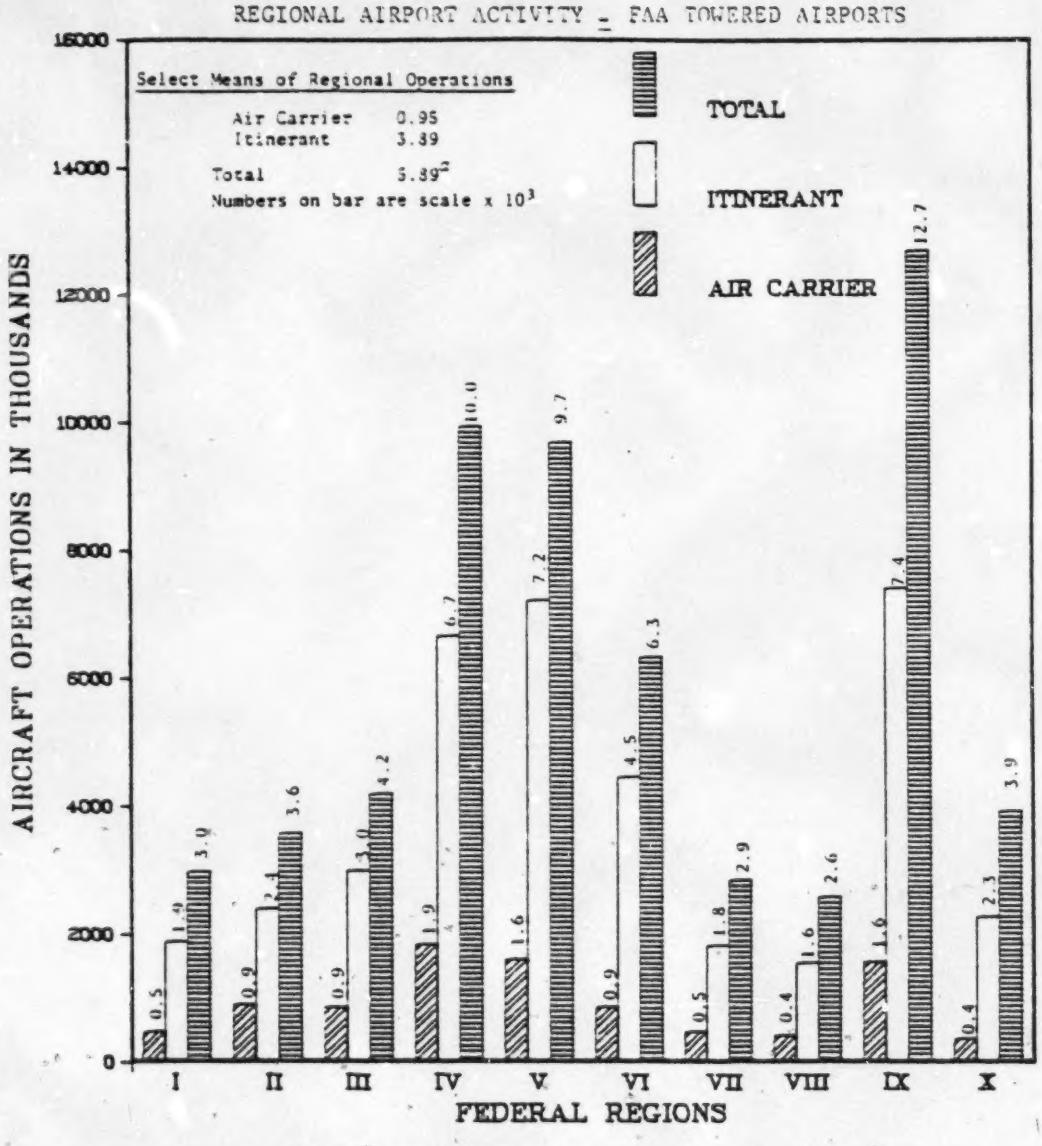


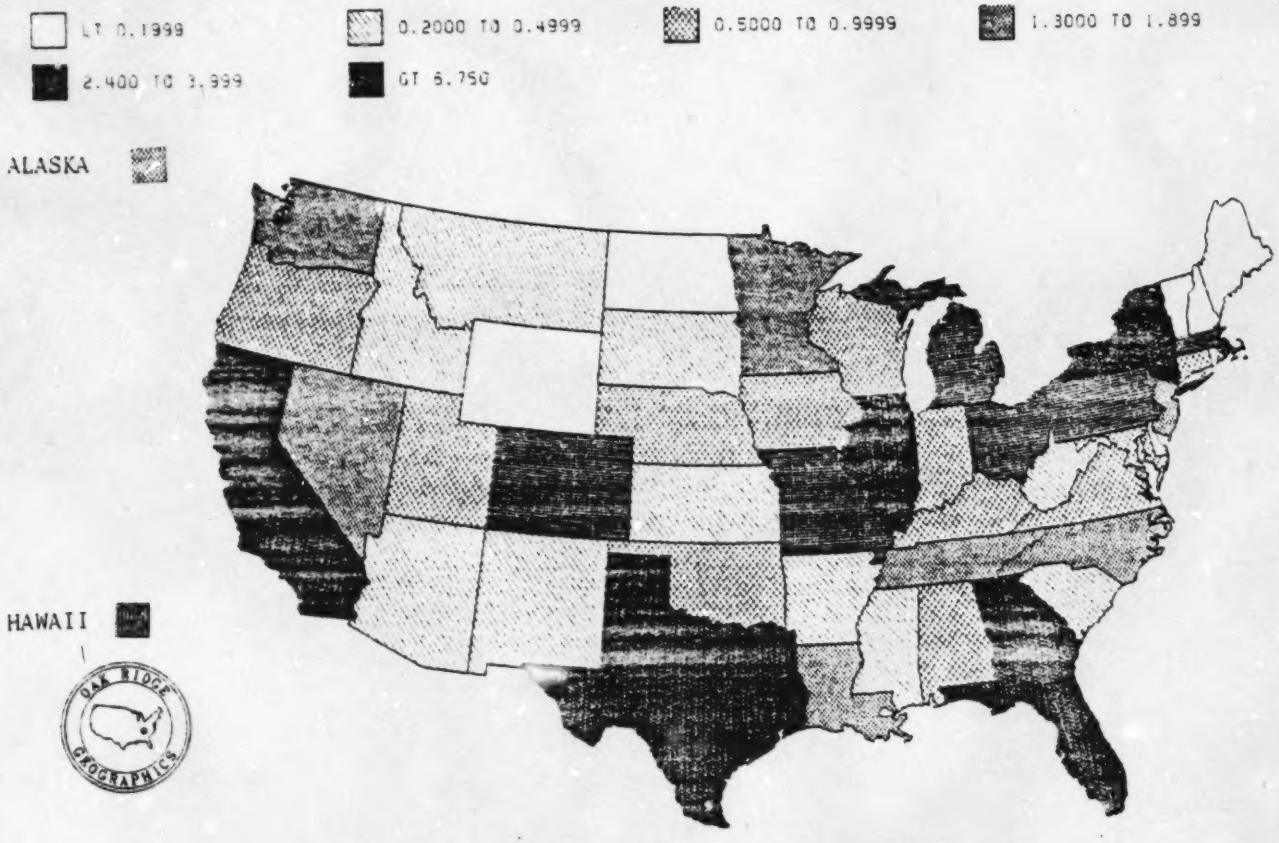
Fig. 2-2. Regional Activity at FAA-Towered Airports FY 1975.

Source: U.S. Department of Transportation, Federal Aviation Administration, Terminal Area Forecasts 1978–1988, Washington, D.C., April 1977.

SAMPLE 5

DATA GROUP: Regional Airport Activity - FAA Towered Airports

ELEMENT	Operations	OF	air carriers	IN	thousands
ELEMENT	Operations	OF	itinerants	IN	thousands
ELEMENT	Operation	OF	total	IN	thousands
BY	FAA-Region	OF	Originating Airport		
FOR	Fiscal year	EQ	1975		



Air Carrier Passenger Enplanements by State, 1975.
(percent of U.S. air carrier passenger enplanements)

Source: Civil Aeronautics Board and U.S. Department of Transportation, Federal Aviation Administration, Airport Activity of the Certificated Route Air Carriers, 12 Months Ended December 31, 1975, Washington, D.C., December 1975.

SAMPLE 6

DATA GROUP: Air Carrier Passenger Enplanements by State

ELEMENT Passenger Enplanements OF Air Carriers IN percent ranges

BY Year EQ 1975

WHERE see domain description for ranges

DOMAIN Passenger Enplanements
TYPE IS RANGES OF Passenger Enplanements %
VALUES ARE 1,2,3,4,5,6
BINDING RULE

1	MEANS	0	THRU	0.1999
2	MEANS	0.2000	THRU	0.4999
3	MEANS	0.5000	THRU	0.9999
4	MEANS	1.3000	THRU	1.8990
5	MEANS	2.4000	THRU	3.9990
6	MEANS	6.7500	THRU	100.000

DOMAIN Passenger Enplanements %
TYPE IS METRIC UNITS Percent of U.S. air carrier passenger enplanements
VALUES ARE 0 THRU 100.0

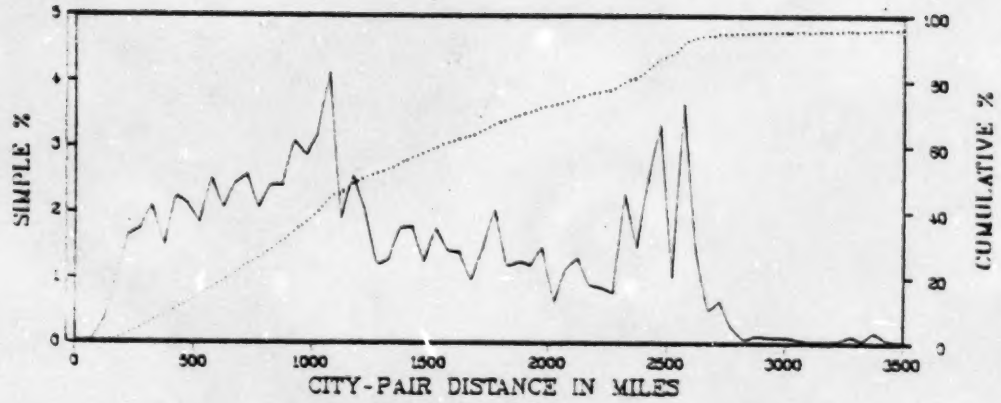


Fig. 2-14. Percent of Passenger Miles by City-Pair Distance[2].

[2]THE PLOTS ARE NOT TO THE SAME SCALE.

Source: P. Gavel, Civil Aeronautics Board, Supplement to the Handbook of Airline Statistics, Washington, D.C., December 1977, pp. 103–104.

SAMPLE 7

DATA GROUP: Percent Passenger Miles by City-Pair Distance

ELEMENT	Passenger Miles	OF	travelled	IN	Simple Percent
ELEMENT	Passenger Miles	OF	travelled	IN	Cumulative Percent
BY	city-pair miles	OF	distance		
FOR	year	EQ	1975		

APPENDIX 2

SYNTAX AND GENERAL RULES FOR LOGICAL DATA DEFINITIONS

SYNTAX AND GENERAL RULES FOR LOGICAL DATA DEFINITIONS

1. Each domain definition has a single TYPE clause which specifies the kinds of meaning expressed by values of the domain. The UNITS phrase is used for expressing the units associated with a metric domain. Typical specifications would include nnn-barrels-per-day, pounds, miles, percent, dollars, etc. The NOMINAL phrase specifies that the values are themselves meaningful or otherwise unit-less. The RANGE and CLASS phrases specify that domain values represent ranges of values or classes of values.

2. The Binding clause specifies the rule for associating legal values with objects identified by the domain.

3. Each domain definition includes one or more phrases of the VALUES clause designating legitimate values for the domain and constraints on those values. The user-specified representation phrase is included to cope with unique representations such as part-number codes, Roman numerals, and hexadecimal numbers. The CODE phrase is used to specify a standard domain coding such as U.S. Postal Service zip code or state abbreviation.

4. If the NULL clause is omitted, the null value is the system-default.

5. The ORDERED clause permits the specification of an ordering of domain values other than the default numeric or alphabetic order. Ordering the domain STATE by two letter postal code could be specified by this clause. The order specified by this clause is used in the THRU phrase of the VALUES clause and the TYPE clause. If the ORDERED clause is omitted, ORDER BY VALUE is assumed.

6. If the CREATION clause is omitted, new values may not be added to the domain.

7. If the DELETION clause is omitted, then values may not be deleted from the domain.

8. If the CREATION or DELETION clauses are specified, then an AUTHORIZATION clause must also be specified.

9. The units specified for a domain named in an ELEMENT or BY or FOR clause must be defined in a UNITS phrase of the corresponding DOMAIN clause.

10. The BY and FOR clauses are identical except for the listing of all the data values assumed by the domains named in the FOR clause.

11. The WHERE clause permits the specification of contextual information not represented by domain values specificable in BY and FOR clauses.

12. The DESCRIPTION clause permits the specification of an extended textual description of the data element.

13. The FIELD...RECORD clause is used to specify in which records values for this data element appear.

APPENDIX 3

LOGICAL DATA DECLARATIONS

LOGICAL DATA DECLARATIONS

DATA BASE DEFINITIONS

DATA BASE NAME IS name [ALIAS IS name] ...

DOMAIN NAME IS name [ALIAS IS name] ...

{TYPE / ENTITY} IS { METRIC [UNITS units spec] / COMPUTED [BY procedure / FROM {domain name}...] / NOMINAL description / {CLASSES / RANGES} OF domain-name }

[{BINDING / DEFINITION}] . [{value MEANS {value [THRU value]} ...}... / RULE IS description]

{VALUES / SYMBOLS} ARE [{value [THRU value]} ... / SUBSET OF domain-name / User-specification / PICTURE picture & spec / CODE code-spec]

```
[NULL IS                    {value       }]
                            {NOT ALLOWED}

[ORDERED BY                 {VALUE                }]
                            {user-specification   }
                            {PROCEDURE procedure  }

[CALL procedure                 ||BEFORE       ||   ||CREATE  ||]
                                ||ON ERROR DURING||   ||RETRIEVE||
                                ||AFTER        ||   ||DELETE  ||

[CREATION IS                {AUTOMATIC   }]
                            {MANUAL      }
                            {NOT ALLOWED }

[DELETION IS                {AUTOMATIC   }]
                            {MANUAL      }
                            {NOT ALLOWED }

  [AUTHORIZATION BY administrator]

DATA GROUP NAME IS name [ALIAS IS name]...
      {ELEMENT      domain [IN units] OF role-specification}...
      [BY           {domain [IN units]}... OF role-specification]...
      [FOR          domain [IN units] OF role-specification
                                EQUALS {value [THRU value]}...  ]...
      [WHERE        general-content-specification]
      [DESCRIPTION  IS text]
      [SOURCE       IS text]
      [FIELD        field-name RECORD record-name].
```

PHYSICAL DECLARATIONS

```
FILE DEFINITIONS NAME IS name [ALIAS IS name] ...
    FOR level-1-name DATA BASE
FILE NAME IS name [ALIAS IS name] ...
    SIZE IS integer PAGES
        [EXPANDABLE          [BY steps OF integer]]
                             [TO integer PAGES   ]
        PAGE SIZE is integer {CHARACTERS}
                             {WORDS     }
    RECORD NAME IS name [ALIAS IS name] ...
        PLACEMENT IS
        {  SYSTEM - DEFINED                                                 }
        {  SEQUENTIAL {field name {ASCENDING }}...                          }
        {                         {DESCENDING}}                             }
        {             DUPLICATES NOT ALLOWED                                }
        {  CALCULATED BY procedure [USING {identifier}...]                  }
        {  DIRECT USING {identifier}...                                     }
        ACCESS IS
        {  [RANGE] KEY {field name {ASCENDING }}...                         }
        {                          {DESCENDING}}                            }
        {  PROCEDURE procedure [USING {field name}...]                      }
        {  SET [LINKED TO [NEXT ]]                                          }
        {                 [PRIOR]                                           }
        {                 [FIRST]                                           }
```

DATA IS

level # { FILLER / field name / DATA }

[REDEFINES field name]

[OCCURRS { { integer / name } TIMES / IF condition }]

[TYPE IS { { { BINARY / DECIMAL } { FIXED / FLOAT } { REAL / COMPLEX } [# ,#] / { BIT / CHARACTER } [#[DEPENDING ON data name]] } }]

[ALIGNMENT IS integer { BITS / CHARACTERS / WORDS }]